Determination of the Stereochemistry of Natural Products from Nuclear Magnetic Resonance Data by Constrained Molecular Dynamics

MICHAEL FALK,* PETER F. SPIERENBURG, and JOHN A. WALTER

Institute for Marine Biosciences, National Research Council of Canada, 1411 Oxford Street, Halifax, Nova Scotia, Canada B3H 3Z1

Received 8 February 1995; accepted 14 June 1995

ABSTRACT

We have developed a molecular modeling procedure to determine the relative configuration of a chiral molecule from nuclear magnetic resonance (NMR) data. Our procedure uses constrained molecular mechanics, and the constraints are interproton distances derived from the experimental nuclear Overhauser enhancement (NOE) data. The main feature is a period of high-temperature dynamics in which frequent inversions occur at most chiral sites. This allows the distance constraints to guide the molecule into configurations consistent with the NOE data. For molecules with complex ring systems, high-temperature dynamics alone may fail to invert certain chiral centers with sufficient frequency. We have countered this by allowing as an option additional inversions of selected chiral centers. The procedure tested successfully on organic molecules of known stereochemistry, with 5 to 17 chiral centers, provided that the number of available constraints was at least twice the number of chiral centers. The procedure is tolerant of large errors in the estimated interproton distances and is reasonably rapid. For a series of sugars, the time required increases less than quadratically with the number of atoms. © 1996 by John Wiley & Sons, Inc.

Introduction

When the molecular structure of a new natural product is determined, an ambiguity often remains about the stereochemistry of its chiral centers. This ambiguity can sometimes be resolved by X-ray crystallography, but if crystals are not available, the configuration of the chiral centers may often be determined from NMR measurements in solution. Many H···H distances and H—C—C—H angles of a molecule can be estimated from nuclear Overhauser enhancements

*Author to whom all correspondence should be addressed.
E-mail: falkm@imb.lan.nrc.ca

Journal of Computational Chemistry, Vol. 17, No. 4, 409–417 (1996)

CCC0192-8651/96/040409-09

(NOEs) and vicinal proton coupling constants, and this is often enough to deduce the stereochemistry. Since the correct structure and its enantiomer yield identical NMR spectra, only the relative configuration of the chiral centers can be deduced from such spectroscopic data.

When a molecule has many chiral centers and some conformational flexibility, very large numbers of possible configurations and conformations have to be explored while monitoring the degree of agreement with the NMR data. In such cases, computer-based molecular modeling is required to find every possible configuration that agrees with the data.

We have developed a procedure, which we call ConGen, that determines the configuration of chiral molecules. ConGen is based on an earlier observation[1] that at 1500 K, chiral centers of molecules subjected to molecular dynamics[2,3] occasionally invert from R to S or from S to R. With most force fields, the probability of such inversions increases with temperature until at 8000 K most chiral centers invert back and forth rapidly. When dynamics is used for conformational searches on molecules of known chirality, such inversions are a nuisance; but for molecules of unknown or uncertain chirality, their occurrence can be put to good use by allowing configurational searching.

ConGen consists of repeated cycles of high-temperature molecular dynamics under distance constraints derived from the NOEs. In the case of complex ring systems, some chiral centers may be only rarely inverted by dynamics, even at 8000 K. In such cases, the dynamics runs have to be supplemented by the inclusion of additional inversions.

ConGen generates a series of possible configurations of the molecule, the correct configuration being identifiable by yielding the best agreement with the NOEs and with other experimental data. Our use of constrained high-temperature dynamics allows us to settle the relative configuration of all chiral centers at once and distinguishes our procedure from that recently proposed by Reggelin et al.,[4] which has been shown to settle the uncertain configuration at one chiral center.

The ConGen Technique

COMPUTING ENVIRONMENT

The work reported in this article was performed using SYBYL molecular modeling software (Version 6.1, Tripos, Inc., St. Louis, MO) running on a Personal Iris 4D/35 TurboGraphic workstation (Silicon Graphics, Inc., Mountain View, CA) under the IRIX 5.2 operating system. The Tripos force field[5] without the electrostatic term was used in molecular dynamics and in minimizations. The time step in dynamics was normally set at 1 fs. Shorter time steps may be necessary if the dynamics runs crash due to the generation of very high velocities. The ConGen script is written as a set of macros in SYBYL programming language and is available upon request (see note at the end of this article).

DISTANCE CONSTRAINTS

The N observed NOE signals (usually the cross-peak intensities in 2D NOE spectra) are converted to N approximate interproton distances D_i by a suitable calibration procedure.[6] Each interproton distance D_i is then used as a distance constraint by adding an energy term E_i to the force field:

$$E_i = \tfrac{1}{2}k(d_i - D_i)^2 \qquad \text{when } d_i > D_i \qquad (1a)$$

$$E_i = 0 \qquad \text{when } d_i \leq D_i \qquad (1b)$$

where d_i is the distance corresponding to D_i in the given model structure and k is a variable force constant. A ConGen constraint table is created; its N rows correspond to the N distance constraints.

THE CONSTRAINED DYNAMICS MODULE

A ConGen run requires an initial structure. To avoid bias, the initial configuration is scrambled at the start of every ConGen run by having each of its chiral centers inverted or not inverted at random.† The randomized structure has its energy minimized without constraints and then again with the constraints applied in full (k = 400 kcal mol^{-1} Å^{-2}). It is then subjected to a plateau period (200 fs) of dynamics at high temperature (8000 K). The dynamics run is continued through the anneal period (800 fs), in which the temperature is lowered in 80 steps from 8000 K to 300 K. The dynamics run is then stopped and the constraint force constant k is reduced in 20 steps to zero while the energy is minimized at every step. The final un-

†The invert atom procedure in SYBYL will not work in many ring situations. Several macros had to be created and included in ConGen to allow every type of chiral site to be inverted at will.

constrained structure is stored in a database, while some of its geometric parameters and its energy are stored in one row of the output table. It is then used as the starting structure in the next cycle.

THE NEED FOR THE INVERSION MODULE

When several consecutive cycles of constrained dynamics return the same configuration, it is usually a sign that the correct structure has been reached. Alternately, however, the molecule might be stuck in a structure that is only partially correct but cannot be further changed by dynamics alone. We have found that two distinct phenomena can be responsible for this:

1. Once a correct structure has been locally achieved, rigid sections of the molecule, which contain several chiral centers (e.g., the fused-ring system in okadaic acid, shown in Fig. 1), tend to retain their configuration in further dynamics cycles.
2. Even at 8000 K, some chiral centers invert at a very low rate. This is likely to occur for spiro junctions and for sterically crowded carbons. A low rate of inversion would be clearly revealed by a run of unconstrained dynamics.

Occurrence of both phenomena may prevent high-temperature dynamics from producing the configuration yielding the best agreement with the distance constraints. As a countermeasure, we have included in ConGen an option to supplement the cycles of dynamics with an inversion module. This module performs two types of operation: multiple-site inversion, which inverts all chiral centers of a selected region of the molecule, and single-site inversion, which inverts one selected chiral center. These two operations have been designed to counter phenomena 1 and 2, respectively. The regions or centers to be inverted are selected by the user at the beginning of the run. For multiple-site inversions, this selection is based on the known molecular connectivity; while for single-site inversions, it may be based on a preliminary run of unconstrained dynamics. (To run ConGen without constraints, one may simply set the constraint force constant k to 0 in the constraint table.)

If the inversion option is accepted, the inversion module is implemented automatically whenever three consecutive cycles of constrained dynamics return the same configuration. It will carry out first the multiple-site inversions and then the single-site inversions that have been selected by the user.

MULTIPLE-SITE INVERSIONS

The regions to be selected for multiple-site inversion should be those that are relatively rigid (e.g., spiro junctions or fused-ring systems). Each region to be inverted must first be isolated from the rest of the molecule by temporarily cutting one, two, or more bonds. For example, in the case of okadaic acid, which possesses three ring systems, the central region containing the chiral carbons C16, C19, C22, C23, C24, C26, and C27 could be selected. Two bonds would have to be chosen for cutting; it could be those marked X in Figure 1. In each multiple-site inversion, the selected bonds are cut, isolating the part of the molecule that is to be inverted. The inversion is performed with SYBYL's invert command, after which the cut bonds are restored. Each of the resulting partially inverted structures (which at this point may be highly strained) is subjected to an energy minimization and then to one cycle of constrained dynamics. The partly inverted structures and the

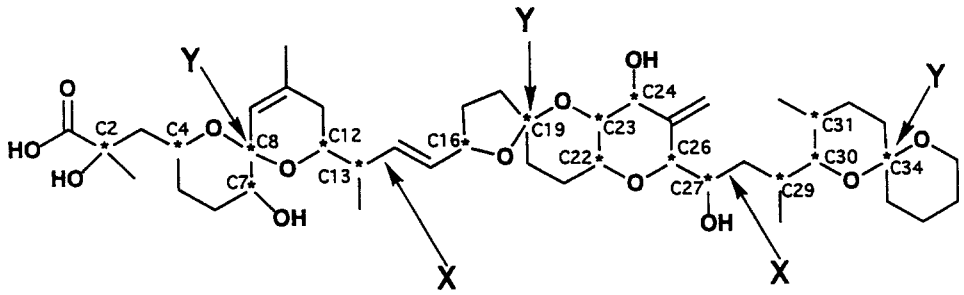

FIGURE 1. Structure of okadaic acid. Chiral carbon atoms are starred and numbered. Bonds marked X are selected to isolate the central region for multiple-site inversion. Atoms marked Y are selected for single-site inversion.

uninverted structure are compared, and the one showing the lowest number of violations is retained. If one of the multiple-site inversions results in an improved structure, the inverted structure is stored in the database and recorded in the output table and becomes the starting structure in the next cycle of restrained dynamics. Otherwise, the uninverted structure is submitted to single-site inversions at selected chiral sites.

SINGLE-SITE INVERSIONS

The atoms to be selected for single-site inversions should be all those which form spiro junctions between two rings, plus any other chiral carbons for which a low rate of inversion has been found in a preliminary run of unconstrained dynamics. For example, in the case of okadaic acid, the selected atoms should be those marked Y in Figure 1. In each single-point inversion, the structure is inverted locally at the selected atom. Each of the resulting structures is subjected to an energy minimization and then to one cycle of constrained dynamics. The inverted structures and the uninverted structure are then compared and the one with the lowest number of constraint violations is stored in the database, recorded in the output table, and chosen as the starting structure in the next cycle of constrained dynamics. A ConGen run continues until the operator, upon examining the output table, decides to stop it. The total flow chart of ConGen is shown in Figure 2.

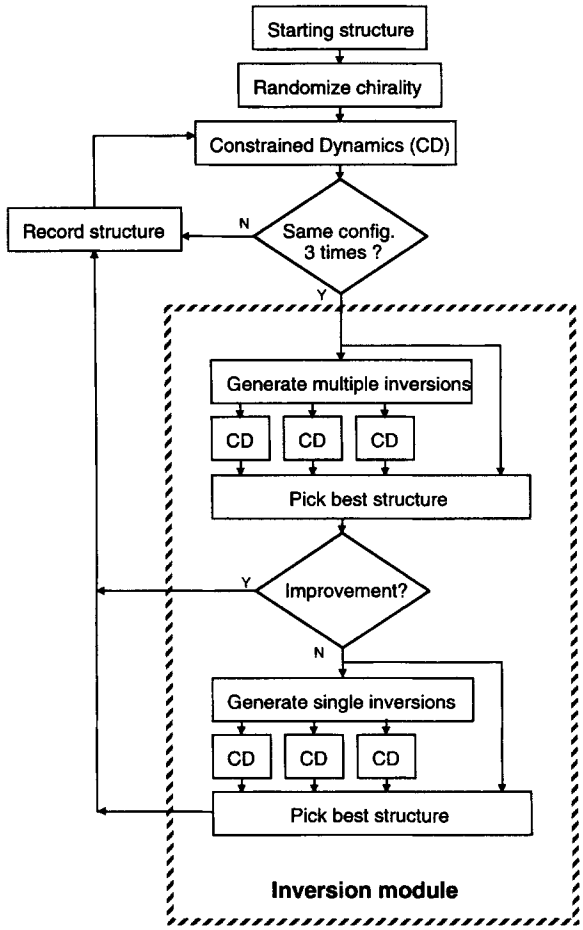

FIGURE 2. Overall flow diagram of ConGen.

ANALYSIS OF THE RESULTS

ConGen produces an output table whose rows correspond to the structures generated by ConGen. The first N columns of the table record the N interproton distances d_i corresponding to the constraints. Additional columns record chirality, energy, and parameters V and σ, both of which measure the agreement of a structure with the set of distance constraints. V is the number of constraints that are being violated by more than 10%, while σ is the root mean square (rms) violation, in Å units, defined by the following equations:

$$\delta_i = (d_i - D_i) \quad \text{when } d_i > D_i \qquad (2a)$$

$$\delta_i = 0 \quad \text{when } d_i \leq D_i \qquad (2b)$$

$$\sigma^2 = (1/N)\sum \delta_i^2 \qquad (2c)$$

where the summation is over the N constraints. To allow for the internal rotation of methyl groups, the constrained distance d_i to a proton of a methyl group is taken to be the shortest of the distances to any one of the three methyl protons. The output table can be examined offline for plausible structures (i.e., those which have low energy and a low value of V and σ; the ideal value of V and σ is 0). Multiple ConGen runs are made, each with a differently scrambled initial starting structure, until it is judged that every plausible structure has been found.

In the most favorable case, ConGen yields only one plausible structure (and its enantiomer), which must then be tested for agreement with all available experimental data. One such test is the agreement of the remaining interproton distances, which were not used as constraints, with the NOEs present (or absent). Another test is the agreement between the values of the H—C—C—H dihedral angles and those calculated by the Karplus equation[7,8] from the vicinal proton coupling constants. In the case that ConGen yields several plausible

structures, these tests may eliminate all but one of them. If several plausible structures survive the testing, additional experimental data are needed to decide between them.

Tests Using Simulated Constraints

We have tested the performance of ConGen in a controlled way on selected molecules, using their known crystallographic structures to derive H ··· H distance constraints. We then varied the number of constraints actually used by ConGen and imposed random errors on the constrained distances, thus simulating experimentally derived constraints. The test molecules included a series of carbohydrate structures[9–12] up to the tetrasaccharide level, containing from 5 to 17 chiral sites. This series, containing one pyranose of furanose ring per sugar unit, provided us with molecules of moderate complexity and allowed us to explore the effect of increasing molecular size on ConGen performance. For an example of a more complex ring system, we chose okadaic acid[13] (Fig. 2). This naturally occurring marine toxin contains 17 chiral sites, three spiro junctions, and a total of seven rings, two of them fused.

Each test molecule had its energy minimized and was then stored as the target structure. We recorded the N interproton distances D_i shorter than 3.5 Å, omitting geminal and vicinal proton pairs as trivial. These distances were used to generate N distance constraints. To determine the minimum number of constraints required for ConGen to succeed, constraints were gradually removed at random. To determine the effect of errors which would occur if the interproton distances were estimated from the NOEs, random error was imposed on the distance constraints by means of an adjustable uncertainty factor F. When $F = 0$ Å^{-1}, the interproton distances D_i used in the constraints are exact. When $F > 0$ Å^{-1}, each D_i is increased by a random amount ΔD_i:

$$\Delta D_i = F \cdot \text{rand} \cdot (D_i - 2.5)^2 \qquad (3)$$

where rand is a random number in the range (0, 1). The form of eq. (3) allows for the tendency of errors in interproton distances derived from NOE data to increase rapidly with the length of such distances.[14] When $D_i = 3.5$ Å, the mean errors thus imposed on the interproton distances are 1.5 Å when $F = 3$ Å^{-1} and 2.5 Å when $F = 5$ Å^{-1}.

For each test molecule, several starting structures were generated by inverting or not inverting its chiral atoms at random, thus yielding structures with chirality very different from the target structure and from one another. The energy was minimized, and 40 to 80 cycles of ConGen were run under the complete set of N distance constraints and $F = 0$. The table of results was then examined to see whether the run had been successful; success was defined as the fulfillment of all three of the following criteria:

1. The target structure or its enantiomer is returned repeatedly, more often than any other structure.
2. The target structure or its enantiomer yields lower values of V and of σ than any other structure.
3. Success is attained with at least three differently scrambled starting structures.

The run was repeated with gradually diminishing N and increasing F until the success zone of the ConGen procedure became defined for each of the test molecules.

Results

All test molecules yielded a success zone at high values of N and low values of F. Within this zone, ConGen eventually returned the correct structure or its enantiomer from every scrambled initial structure. Within the success zone, once the correct structure was returned, further cycles of ConGen continued to return it a large proportion of the time. At low values of N and high values of F lay the failure zone, in which the correct structure was either never returned or was returned sporadically among many incorrect structures. The two zones were separated by an intermediate band, in which ConGen runs sometimes succeeded and sometimes failed.

Within the success zone, a typical ConGen output table looks like the example in Table I, from a run on okadaic acid with $N = 60$ and $F = 0$. In the output table, the chirality of each structure returned by ConGen is represented by a string of 17 R of S symbols (grouped into threes for convenience), corresponding to the 17 chiral carbons in standard order of numbering: C2 C4 C7, C8 C12 C13, C16 C19 C22, C23 C24 C26, C27 C29 C30, C31 C34.

TABLE I.
Typical Progress of a ConGen Run on Okadaic Acid.[a]

CYCLE	CHIRALITY	E	V	σ
0	R|R|R S|S|S S S|R S|S|S |R|S|R S|S	-	-	-
1	S R S S R S S|R R S R S S|R|S |S|S	15.62	14	0.70
2	S R S S R S S|R R S R S S|R|S |S|S	16.04	13	0.77
3	S R S S R S S|R R S R S S|R|S |S|S	17.08	15	0.78
4	S R S S R S S S S R S R R|S|R |R R	12.97	7	0.39
5	S R|R S R S S S S R S R R|S|R |R R	15.26	9	0.42
6	S R S S R S S S S R S R R|S|R |R R	9.31	4	0.37
7	S R S S R S S S S R S R R|S|R |R R	9.65	5	0.36
8	S R S S R S S S S R S R R|S|R |R R	9.37	7	0.38
9	S R S S R S S S S R S R R R|S |S|S	8.99	4	0.31
10	S R S S R S S S S R S R R R|S |S|S	11.33	6	0.33
11	S R S S R S S S S R S R R R|S |S|S	7.28	7	0.52
12	S R S S R S S S S R S R R R|S |S|S	10.88	9	0.36
13	S R S S R S S S S R S R R R R S R	21.05	3	0.29
14	S R S S R S S S S R S R R R|S |S|S	5.74	7	0.57
15	S R S S R S S S S R S R R R|S |S|S	14.16	8	0.37
16	S R S S R S S S S R S R R R|S |S|S	9.42	5	0.33
17	S R S S R S S S S R S R R R|S |S|S	9.42	5	0.33
18	S R S S R S S S S R S R R R|S |S|S	9.67	7	0.42
19	S R S S R S S S S R S R R|S S R S	6.55	2	0.10
20	S R S S R S S S S R S R R|S S R S	7.57	2	0.10
21	S R S S R S S S S R S R R|S S R S	6.69	2	0.10
22	S R S S R S S S S R S R R R R S R	6.65	3	0.18
23	S R S S R S S S S R S R R R R S R	8.57	4	0.24
24	S R S S R S S S S R S R R R R S R	4.89	0	0.04

[a] E (kcal / mol) is energy, V is the number of constraints that are violated by more than 10%, and σ (Å) is the rms violation index defined by eq. (2c). Only the first 24 of the 80 structures returned are shown. Cycle 0 is the starting structure. Structures with the correct chirality were returned in cycles 13, 22–41, 49–56, 59–66, and 70–80.

Since spectroscopic results cannot distinguish between a structure and its enantiomer, the enantiomeric structure SRS SRS SSS RSR RRR SR and the actual X-ray structure RSR RSR RRR SRS SSS RS must be considered equally correct. Clearly, an R or S in a given site is inherently neither correct nor incorrect. Correctness applies to pairs of consecutive sites: If a particular pair in the X-ray structure is RS, then RS or SR are both correct while SS and RR are not. The incorrect pairs, or "faults," along each chiral string are marked by vertical lines | in Table I. The starting structure of the run depicted in Table I had 12 faults. The first ConGen cycle diminished the number of faults to five, and further cycles gradually eliminated faults until eventually a structure with totally correct configuration appeared in cycle 13. It disappeared in cycle 14 but returned in cycle 22, and after that it persisted from cycle to cycle, about 75% of the time. At its first appearance, the correct configuration was accompanied by a conformation that was grossly incorrect, as shown by its relatively high energy and by large values of V and σ. The conformation gradually improved, shown by a lowering of energy and a decrease in V and σ (which, however, only rarely attained their ideal value of zero). Thus, while ConGen arrives at the correct configuration, it will generally not return it every time, and it will only occasionally return it with the correct conformation. This is not surprising, considering the large number of low-energy conformations available for any one configuration of a molecule as complex as okadaic acid.

The minimum number of distance constraints required for success was typically between two and three times the number of chiral centers. ConGen was found to be tolerant of random error in the interproton distances used in the constraints, typically up to factor $F = 3$ Å^{-1} or higher, corresponding to a mean error of 1.5 Å for an interproton distance of 3.5 Å.

For the series of carbohydrate molecules, cycles of constrained dynamics led to the correct configuration without the need to use the inversion mod-

TABLE II.
Performance of ConGen with Test Molecules

Molecule	Glucose $C_6H_{12}O_6$	Sucrose $C_{12}H_{22}O_{11}$	1-Kestose $C_{18}H_{32}O_{16}$	Nystose $C_{24}H_{42}O_{21}$	Okadaic Acid $C_{44}H_{68}O_{13}$
Crystal structure	ref. 10	ref. 11	ref. 12	ref. 13	ref. 14
Chiral centers	5	9	13	17	17
Available constraints	13	27	46	54	103
Minimum constraints needed	9	24	28	36	45
Maximum tolerable F ($Å^{-1}$)	5	4	5	4	4

ule. For okadaic acid, however, constrained dynamics repeatedly returned structures containing a portion of correct chirality, with the remaining portion consisting of several chiral centers and being the enantiomer of the correct structure. This problem was solved through the intervention of the multiple-site and single-site inversions.

An overview of the performance data for different test molecules is shown in Table II. The average times required to complete one cycle of constrained dynamics for the four oligosaccharides tested, and for two larger oligosaccharides, are shown in Table III. It appears that the time per cycle increases approximately as $A^{1.7}$, where A is the number of atoms. This compares favorably with many other search methods. For example, distance geometry algorithms,[15] which are an alternate means of searching configuration space,[16] require bound smoothing routines[15] that nominally scale as A^3.

Test Using Experimental Constraints

The bulk of our tests on molecules of known structure were performed using simulated constraints derived from known interhydrogen distances. We have also tested ConGen using experimentally derived constraints. We chose the antitumor agent taxol, $C_{47}H_{51}NO_{14}$, whose stereochemistry is well known[17] and for which we recorded a single NOESY spectrum in $CDCl_3$, with a mixing time of 400 ms, using a Bruker AMX 500 spectrometer. We were mainly concerned with the relatively rigid central core region of taxol, which contains a complex set of fused four-, six-, and eight-member rings (see Fig. 1 of ref. 17). This region has nine chiral atoms: C1(S), C2(S), C3(R), C4(S), C5(R), C7(S), C8(S), C10(R), and C13(S) (there are also two chiral atoms in the taxol sidechain). Using our chirality string representation, the correct chiral structure of the core region is therefore SSR SRS SRS. The $N = 39$ observed NOE cross peaks that belonged to this region were classified as very weak, weak, medium, strong, or very strong and were used to generate a total of 39 distance constraints with the interproton distance D_i set, respectively, to 3.3, 3.1, 2.9, 2.7, or 2.5 Å on the basis of experience.

The starting structures for ConGen were generated by constructing the taxol molecule on the basis of Figure 1 of ref. 17 and then randomly inverting its chiral centers in the usual manner. We ran ConGen using single-site inversions at C4

TABLE III.
Average Time Taken by Silicon Graphics 4D / 35 Workstation per Cycle of Constrained Dynamics

Saccharide	Mono	Di	Tri	Tetra	Hexa	Octa
Molecular Formula	$C_6H_{12}O_6$	$C_{12}H_{22}O_{11}$	$C_{18}H_{32}O_{16}$	$C_{24}H_{42}O_{21}$	$C_{36}H_{62}O_{31}$	$C_{48}H_{82}O_{41}$
No. of chirals	5	9	13	17	25	33
No. of atoms	24	45	66	87	129	171
Time (sec)	47	100	223	332	682	1072

and C8 (low rates of inversion had been observed for these chiral atoms in a preliminary run) and no multiple-site inversions. ConGen runs using sets of 39, 27, 24, or 19 constraints yielded repeatedly and almost exclusively the correct core configuration SSR SRS SRS or, in about half the runs, its enantiomer RRS RSR RSR. The configuration of the two chiral sites on the sidechain, a region of the molecule which had no constraints associated with it, returned SS, SR, RS, or RR, each about a quarter of the time, as would be expected due to chance. When 19 or 24 constraints were used, several cycles returned incorrect core configurations before the correct configuration was first returned; but when 27 or 39 constraints were used, the correct core configuration was typically returned already by the first ConGen cycle. We have found that, in general, the minimum computing time required to ascertain the correct configuration by a ConGen run decreases rapidly when additional constraints are added. Success of ConGen with taxol, using experimental NOEs, is especially satisfying in view of this molecule's complex set of fused rings.

Analogy to Genetic Algorithms

Genetic Algorithms (GAs)[18,19] have recently shown great promise in conformational searching[20] and in molecular modeling in general.[21] We have noted interesting analogies between some of the procedures we have developed and those used in genetic algorithms: (1) Our chirality strings are analogous to the bitstrings representing candidate solutions in GA, except that ConGen operates on one string at a time, while GA searches operate on populations of bitstrings; (2) our one-site inversion of a chirality string is exactly equivalent to the mutation operation on a bitstring in GA; and (3) our multiple-site inversion closely resembles the one-point and two-point recombination operations in GA, except that recombinations in GA involve pairs of bitstrings, while multiple-site inversions in ConGen involve one chirality string. It is likely that the preceding analogies could be usefully exploited in further development and automation of ConGen using GA methodology along the lines described by Blommers et al.[22]

Conclusions

Searching on the order of 2^{17} configurations, each of them able to assume many conformations, would be too computationally demanding for most systematic or stochastic search methods.[23] However, constrained dynamics has the ability to guide the search rapidly toward the correct configuration, if enough distance constraints are available. The accuracy of these constraints need not be high. The ConGen procedure is efficient in terms of computer time and, for the molecules tested, scales as less than the square of the number of atoms. With the use of multiple- and single-site inversions, the correct configuration (or its enantiomer) was returned for every molecule tested. ConGen should work successfully on molecules of known connectivity, which occur in solution in one main conformation, provided that the molecule contains enough protons in close proximity, with well-resolved resonances, to generate sufficient constraints from NOE data. Where data are insufficient, ConGen might be used to decide the relative stereochemistry of portions of the molecule, or to decide what extra data would be needed to complete the description. We are now in the process of applying ConGen to several new natural products isolated in this laboratory.

Note

ConGen script, in SYBYL programming language (SPL), is available from the authors upon request, along with the mol2 file of a test molecule, a constraint table, an output table, and a typical log file. Arrangements are also being made with Tripos, Inc. to place the script in the SPL Library distributed with SYBYL and, in addition, to make it accessible through the Tripos Home Page on the World Wide Web.

Acknowledgments

We thank James A. Sedgwick for programming and useful suggestions; Dr. Carlos Lucasius of Dalhousie University for discussions on genetic algorithms and a critical reading of the manuscript; TPL Phytogen Ltd., Delta, British Columbia, for a sample of taxol; Ian W. Burton for recording the NOESY spectrum; and the Software Support Group at Tripos, Inc., especially Chris van Dyke, for much help throughout this work. This article is NRCC no. 38087. Parts of this work were presented at the Tripos Users' Group Meetings, St. Louis, Missouri, May 26–27, 1994 and May 24–26, 1995, at the 40th Canadian Spectroscopy Conference, Halifax, Nova

Scotia, August 8–10, 1994, and at the 36th Experimental Nuclear Magnetic Resonance Conference, Boston, Massachusetts, March 26–30, 1995.

References

1. J. A. Walter, M. Falk, and J. L. C. Wright, *Can. J. Chem.*, **72**, 430 (1994).
2. R. M. Scheek, W. F. van Gunsteren, and R. Kaptein, *Meth. Enzymol.*, **177**, 204 (1989).
3. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak, *J. Chem. Phys.*, **81**, 3684 (1984).
4. M. Reggelin, H. Hoffmann, M. Kock, and D. F. Mierke, *J. Am. Chem. Soc.*, **114**, 3272 (1992).
5. M. Clark, R. D. Cramer III, and N. Van Opdenbosch, *J. Comp. Chem.*, **10**, 982 (1989).
6. R. Kaptein, R. Boelens, R. M. Scheek, and W. F. van Gunsteren, *Biochemistry*, **27**, 5389 (1988).
7. M. Karplus, *J. Am. Chem. Soc.*, **95**, 2870 (1963).
8. C. A. G. Haasnoot, F. A. A. M. de Leeuw, and C. Altona, *Tetrahedron*, **36**, 2783 (1981).
9. G. M. Brown and H. A. Levy, *Science*, **147**, 1038 (1965).
10. G. M. Brown and H. A. Levy, *Acta Cryst.*, **B29**, 790 (1973).
11. G. A. Jeffrey and Y. J. Park, *Acta Cryst.*, **B28**, 257 (1972).
12. G. A. Jeffrey and D. B. Huang, *Carbohydr. Res.*, **247**, 37 (1993).
13. K. Tachibana, P. J. Scheuer, Y. Tsukitani, H. Kikuchi, D. Van Engen, J. Clardy, Y. Gopichand, and F. J. Schmitz, *J. Am. Chem. Soc.*, **103**, 2469 (1981). The text of this article describes the chirality of C19 of okadaic acid as S, but the atomic coordinates, published as supplementary material to this article, clearly imply that the chirality at C19 is R.
14. J. D. Baleja, J. Moult, and B. D. Sykes, *J. Magn. Res.*, **87**, 375 (1990).
15. I. D. Kuntz, J. F. Thomason, and C. M. Oshiro, *Meth. Enzymol.*, **177**, 159 (1989).
16. M. Reggelin, M. Kock, K. Conde-Frieboes, and D. Mierke, *Angew. Chem. Int. Ed. Engl.*, **33**, 753 (1994).
17. G. N. Chmurny, B. D. Hilton, S. Brobst, S. A. Look, K. M. Witherup, and J. A. Beutler, *J. Nat. Prod.*, **55**, 414 (1992).
18. L. Davis (editor), *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York 1991.
19. C. B. Lucasius and G. Kateman, *Chemometrics Intell. Lab. Systems*, **25**, 1, 99 (1994).
20. R. S. Judson, E. P. Jaeger, A. M. Treasurywala, and M. L. Peterson, *J. Comp. Chem.*, **14**, 1407 (1993).
21. T. Brodmeier and E. Pretsch, *J. Comp. Chem.*, **15**, 588 (1994).
22. M. J. J. Blommers, C. B. Lucasius, G. Kateman, and R. Kaptein, *Biopolymers*, **32**, 45 (1992).
23. M. Saunders, K. N. Houk, Y.-D. Wu, W. C. Still, M. Lipton, G. Chang, and W. C. Guida, *J. Am. Chem. Soc.*, **112**, 1419 (1990).